

MODELOS DE ELECCIÓN DISCRETA

Eva Medina Moral - www.eva.medinaam.es

(Diciembre 2003)

1. INTRODUCCIÓN

2. INTERPRETACIÓN ESTRUCTURAL DE LOS MODELOS DE ELECCIÓN DISCRETA

3. MODELO LINEAL DE PROBABILIDAD (MLP)

Especificación e interpretación del MLP

Limitaciones de la estimación por MCO

4. MODELOS DE PROBABILIDAD NO LINEAL

Especificación de los modelos de elección discreta (Logit y Probit)

Estimación de los parámetros en los modelos Logit

A. Estimación con observaciones no repetidas: Método de Máxima-Verosimilitud

B. Estimación con observaciones repetidas: Método Mínimos Cuadrados Generalizados

Contraste y validación de hipótesis

A. Significatividad estadística de los parámetros estimados

B. Medidas de bondad de ajuste del modelo

5. MODELOS DE RESPUESTA MÚLTIPLE

Modelos de respuesta múltiple con datos no ordenados

A. El modelo Logit Multinomial

B. El modelo Logit Condicional

C. El modelo Logit Anidado

Modelos de respuesta múltiple con datos ordenados

1. INTRODUCCIÓN

La utilidad de los modelos de elección discreta frente a la econometría tradicional radica en que los primeros permiten la modelización de variables cualitativas, a través del uso de técnicas propias de las variables discretas. Se dice que una variable es discreta cuando está formada por un número finito de alternativas que miden cualidades. Esta característica exige la codificación como paso previo a la modelización, proceso por el cual las alternativas de las variables se transforman en códigos o valores cuánticos, susceptibles de ser modelizados utilizando técnicas econométricas.

La modelización de este tipo de variables se conoce genéricamente con el nombre de modelos de elección discreta, dentro de la cual existe una amplia tipología de modelos. En concreto, según el número de alternativas incluidas en la variable endógena, se distinguen los modelos de respuesta dicotómica frente a los denominados modelos de respuesta o elección múltiple. Según la función utilizada para la estimación de la probabilidad existe el modelo de probabilidad lineal truncado, el modelo Logit y el modelo Probit. Según que las alternativas de la variable endógena sean excluyentes o incorporen información ordinal se distingue entre los modelos con datos no ordenados y los modelos con datos ordenados. Dentro de los primeros, según que los regresores hagan referencia a aspectos específicos de la muestra o de las alternativas entre las que se ha de elegir, se distingue entre los modelos multinomiales y los condicionales.

Teniendo en cuenta todos los elementos que influyen en el proceso de especificación de los modelos de elección discreta, se puede establecer una clasificación general de los mismos, que queda recogida en la el siguiente cuadro.

Clasificación de los modelos de elección discreta

Nº de alternativas	Tipo de alternativas	Tipo de función	El regresor se refiere a:	
			Características (de los individuos)	Atributos (de las alternativas)
Modelos de respuesta dicotómica (2 alternativas)	Complementarias	Lineal	Modelo de Probabilidad Lineal Truncado	
		Logística	Modelo Logit	
		Normal tipificada	Modelo Probit	
Modelos de respuesta múltiple (más de 2 alternativas)	No ordenadas	Logística	Logit Multinomial	Logit Condicional
			- Logit Anidado - Logit Mixto	- Logit Anidado - Logit Mixto
	Ordenadas	Normal tipificada	Probit Multinomial	Probit Condicional
			Probit Multivariante	Probit Multivariante
Ordenadas	Logística	Logit Ordenado		
	Normal tipificada	Probit Ordenado		

2. INTERPRETACIÓN ESTRUCTURAL DE LOS MODELOS DE ELECCIÓN DISCRETA

En la literatura existen dos enfoques para la interpretación estructural de los modelos de elección discreta. El primero hace referencia a la modelización de una variable latente a través de una función índice, que trata de modelizar una variable inobservable o latente. El segundo de los enfoques permite interpretar los modelos de elección discreta bajo la teoría de la utilidad aleatoria, de tal manera que la alternativa seleccionada en cada caso será aquella que maximice la utilidad esperada.

Para entender ambos enfoques, el razonamiento empleado se aplicará al caso sencillo de la modelización de una variable dicotómica, siendo la aplicación generalizada al caso de los modelos de respuesta múltiple inmediata. Bajo el primero de los enfoques se trata de modelizar una variable índice, inobservable o latente no limitada en su rango de variación, I_i^* . Cuando la variable latente supera un determinado nivel, la variable discreta toma el valor 1, y si no lo supera toma el valor 0. La variable latente depende de un conjunto de variables explicativas¹ que generan las alternativas que se dan en la realidad y que permiten expresar el modelo dicotómico como:

$$Y_i = \begin{cases} 1 & \text{si } I_i^* > 0 \quad \text{lo que ocurre cuando } X_i \mathbf{b} + \mathbf{e}_i > 0 \\ 0 & \text{si } I_i^* < 0 \quad \text{lo que ocurre cuando } X_i \mathbf{b} + \mathbf{e}_i < 0 \end{cases}$$

donde el supuesto sobre la distribución de \mathbf{e}_i determina el tipo de modelo a estimar: si se supone una función de distribución uniforme, se utiliza el Modelo Lineal de Probabilidad truncado; si se distribuye como una normal con media cero y varianza uno, el modelo generado será un Probit; mientras que si se supone que se distribuye como una curva logística, se trataría de un modelo Logit. La hipótesis de que el umbral a superar por la variable latente sea cero se puede modificar por cualquier otro valor sugiriéndose, en determinados estudios, que el valor crítico sea el definido por el término constante.

Bajo este enfoque el modelo probabilístico quedaría definido por

$$P_i = \Pr ob(Y_i = 1) = \Pr ob(I_i^* > 0) = \Pr ob(X_i \mathbf{b} + \mathbf{e}_i > 0) = F(X_i \mathbf{b}) \quad (1)$$

¹ La variable latente está relacionada con sus características a través de un modelo de regresión:

$$I_i^* = X_i \mathbf{b} + \mathbf{e}_i$$

donde $X_i \mathbf{b}$ recibe el nombre de función índice.

Con el modelo así definido, la variable endógena del modelo dicotómico representa la probabilidad de ocurrencia del fenómeno analizado, siendo la probabilidad de que ocurra la opción 1 más elevada cuando mayor sea el valor de I_i^* .

El segundo de los enfoques para la interpretación de los modelos de respuesta dicotómica es el que hace referencia a la modelización a través de la formulación de una utilidad aleatoria. Bajo este enfoque un individuo debe adoptar una decisión que le permita elegir entre dos alternativas excluyentes, la 1 o la 0, lo que hará maximizando la utilidad esperada que le proporciona cada una de las alternativas posibles sobre las que tiene que decidir. Es decir, el individuo i -ésimo elegirá una de las dos alternativas dependiendo de que la utilidad que le proporciona dicha decisión sea superior a la que le proporciona su complementaria.

La formulación del modelo bajo esta teoría parte del supuesto de que la utilidad derivada de una elección, U_{i0} o U_{i1} , es función de las variables explicativas de dicha decisión, que son las características propias de cada una de las alternativas de elección y las características personales propias del individuo, de manera que suponiendo linealidad en las funciones, se tiene

$$\begin{aligned} U_{i0} &= \mathbf{a}_0 + X_{i0} \mathbf{b} + \mathbf{e}_{i0} \\ U_{i1} &= \mathbf{a}_1 + X_{i1} \mathbf{b} + \mathbf{e}_{i1} \end{aligned} \quad (2)$$

donde los \mathbf{e}_{ij} recogen las desviaciones que los agentes tienen respecto a lo que sería el comportamiento del agente medio y que se debe a factores aleatorios. El agente i elegirá la opción 1 si la utilidad de esa decisión supera la de la opción 0 y viceversa, de manera que,

$$Y_i = \begin{cases} 1 & \text{si } U_{i1} > U_{i0} \\ 0 & \text{si } U_{i1} < U_{i0} \end{cases}$$

Y el modelo dicotómico quedaría definido por,

$$\Pr ob(Y_i = 1) = \Pr ob(U_{i1} > U_{i0}) = \Pr ob(U_{i1} - U_{i0} > 0) = F(X_i \mathbf{b}) \quad (3)$$

Según que la función asociada a la perturbación aleatoria \mathbf{e}_{ij} (que será la función de distribución, $F(X_i \mathbf{b})$, que se suponga siga dicha probabilidad), sea una función de distribución uniforme, la función de distribución de la normal tipificada o la de la curva logística, se obtienen el Modelo Lineal de Probabilidad Truncado, el Probit o el Logit, respectivamente.

Ambos enfoques, el de la función índice y el de la formulación de una utilidad aleatoria, justifican en términos estructurales la existencia de los modelos probabilísticos bajo dos teorías económicas alternativas, aunque en ambos casos, la expresión final que define la formulación del modelo es la misma.

3. MODELO LINEAL DE PROBABILIDAD (MLP)

Especificación e interpretación del MLP

La primera tentativa teórica desarrollada para estudiar modelos con variables dicotómicas se planteó como una mera extensión del Modelo Lineal General que viene expresado por:

$$Y_i = a + b_k X_{ki} + e_i \quad (4)$$

donde:

$$Y_i = \begin{cases} 1 & \text{si ocurre una alternativa} \\ 0 & \text{en caso contrario} \end{cases}$$

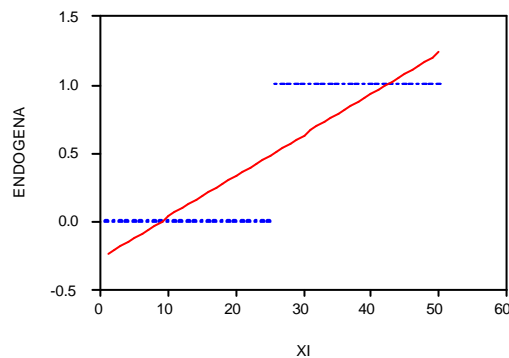
$$X_{ki} = \text{Variables explicativas}$$

$$e_i = \text{Variable aleatoria que se distribuye normal } N(0, \sigma^2)$$

La distribución de la muestra en este tipo de modelos se caracteriza por configurar una nube de puntos de tal manera que las observaciones muestrales se dividen en dos subgrupos. Uno de ellos está formado por las observaciones en las que ocurrió el acontecimiento objeto de estudio ($Y_i = 1$), y el otro, por los puntos muestrales en los que no ocurrió ($Y_i = 0$).

Una representación gráfica de la nube de puntos para el caso de una sola variable explicativa es la que aparece en el gráfico, en el que la variable explicativa se representa en el eje de abscisas y la variable endógena en el eje de ordenadas. La elaboración del modelo lineal de probabilidad requiere el ajuste de esa nube de puntos a una función lineal (recta de regresión) capaz de explicar de la mejor manera el comportamiento de la muestra.

Nube de puntos en un modelo de respuesta dicotómica



El Modelo Lineal de Probabilidad, ecuación (4), se puede interpretar en términos probabilísticos, en el sentido de que un valor concreto de la recta de regresión mide la probabilidad de que ocurra el acontecimiento objeto de estudio. Es decir, \hat{Y}_i se puede considerar como la estimación de la probabilidad de que ocurra el acontecimiento objeto de estudio ($Y_i=1$) siguiendo el siguiente criterio: valores próximos a cero se corresponden con una baja probabilidad de ocurrencia del acontecimiento analizado (menor cuanto más próximos a cero); mientras que a valores próximos a uno se les asigna una probabilidad elevada de ocurrencia (mayor cuanto más próximos a uno).

La interpretación de los coeficientes estimados en los Modelos Lineales de Probabilidad (MLP) es la misma que la del Modelo Lineal General, recogiendo el valor del parámetro el efecto de una variación unitaria en cada una de las variables explicativas sobre la probabilidad de ocurrencia del acontecimiento objeto de estudio. Así, si se produce un incremento de una unidad en la variable explicativa X_{ki} , ese aumento provocaría una variación igual a b_k en la probabilidad $f_i(1)$.

Limitaciones de la estimación por MCO

La estimación del modelo anterior por Mínimos Cuadrados Ordinarios plantea una serie de limitaciones que se pasan a comentar a continuación:

1. El valor estimado puede estar fuera del rango (0 - 1). La estimación del Modelo Lineal de Probabilidad a través de MCO no garantiza que los valores estimados de Y_i estén entre 0 y 1, lo cual carece de lógica al interpretarse el valor estimado como una probabilidad. Este problema se soluciona truncando el rango de variación del valor estimado, dando lugar al modelo conocido con el nombre de **Modelo Probabilístico Lineal Truncado**, y que, para una única variable explicativa, se expresa de la forma:

$$Y_i \quad \begin{cases} 1 & \mathbf{a} + \mathbf{b}_k X_{ki} \geq 1 \\ \mathbf{a} + \mathbf{b}_k X_{ki} & 0 < \mathbf{a} + \mathbf{b}_k X_{ki} < 1 \\ 0 & \mathbf{a} + \mathbf{b}_k X_{ki} \leq 0 \end{cases}$$

Sin embargo, si se restringen los valores de Y_i a 0 y 1, los valores del término independiente y la pendiente varían según los valores de X_i , de tal forma que:

- Para $X_i \leq -\mathbf{a} / \mathbf{b}$: Término independiente y pendiente iguales a 0.
- Para $-\mathbf{a} / \mathbf{b} \leq X_i \leq (1-\mathbf{a}) / \mathbf{b}$: Término independiente igual \mathbf{a} y pendiente igual a \mathbf{b} .

- Para $X_i \geq (1 - \mathbf{a}) / \mathbf{b}$: Término independiente igual a 1 y pendiente igual a 0.

Esto hará que si se incluyen en la estimación puntos en los que $X_i \leq -\mathbf{a} / \mathbf{b}$ ó $X_i \geq (1 - \mathbf{a}) / \mathbf{b}$ los estimadores serán sesgados e inconsistentes.

2. La perturbación aleatoria puede no seguir una distribución Normal: Dados los valores que toma la perturbación aleatoria no se puede asegurar que ésta se distribuya como una normal, al tratarse de una distribución binaria o dicotómica. Si bien el incumplimiento de la hipótesis de normalidad no invalida la estimación por MCO, ya que los estimadores así estimados siguen siendo ELIO, sin embargo, la ausencia de normalidad imposibilita el uso de los estadísticos habituales utilizados para realizar el contraste de hipótesis tales como la t-Student, la F-Snedecor, etc, al basarse dichos contrastes en la hipótesis de normalidad de la perturbación aleatoria.

3. Problemas de heterocedasticidad: Aún en el caso de que se cumplieren las hipótesis de media y correlación nula en la perturbación aleatoria ($E(\mathbf{e}_i) = 0$ y $E(\mathbf{e}_i \mathbf{e}_j) = 0$ para todo $i \neq j$) no se cumple la hipótesis de varianza constante, es decir, la perturbación aleatoria no es homocedástica. Para comprobarlo se calcula la varianza de la perturbación aleatoria a través de su definición:

$$\begin{aligned} \text{Var}(\mathbf{e}_i) &= E(\mathbf{e}_i - E(\mathbf{e}_i))^2 = E(\mathbf{e}_i)^2 = (1 - \mathbf{a} - \mathbf{b}_k X_{ki})^2 f_i(1) + (-\mathbf{a} - \mathbf{b} X_{ki})^2 (1 - f_i(1)) = \\ &= (1 - f_i(1))^2 f_i(1) + (f_i(1))^2 (1 - f_i(1)) = (1 - f_i(1)) f_i(1) (1 - f_i(1) + f_i(1)) = (1 - f_i(1)) f_i(1) \end{aligned} \quad (5)$$

La varianza de la perturbación aleatoria es una función de la probabilidad $f_i(1)$, la cual es a su vez función de cada una de las observaciones de las variables explicativas X_{ki} . La perturbación aleatoria es, por tanto, heterocedástica y la estimación del modelo mediante el método de MCO obtiene unos estimadores de los coeficientes de regresión con varianza no mínima, es decir, no eficientes.

Este problema podría solucionarse estimando el modelo a través de Mínimos Cuadrados Generalizados (MCG). A este tipo de modelos se les denomina Modelos Lineales Probabilísticos Ponderados. La estimación a través de MCG requiere la realización de los siguientes pasos:

- Se estima el modelo (4) mediante MCO sin tener en cuenta el problema de heterocedasticidad, obteniéndose el valor estimado \hat{Y}_i .
- El valor \hat{Y}_i se utiliza para calcular la varianza de la perturbación aleatoria, a través de la fórmula anteriormente obtenida:

$$\text{Var}(\mathbf{e}_i) = (1 - f_i(1))f_i(1) = \hat{Y}_i(1 - \hat{Y}_i) = \mathbf{s}_i^2 \quad (6)$$

- Si los valores estimados de \hat{Y}_i son mayores que la unidad o menores que cero, deben sustituirse por la unidad (en el primer caso) o por cero (en el segundo). En ambos casos el valor resultante del cálculo de la varianza de \mathbf{e}_i será cero, lo que generaría problemas al utilizar la $\text{Var}(\mathbf{e}_i)$ como ponderador. Ante esta situación se puede optar por eliminar las observaciones que generan estos valores, incurriendo en pérdida de información. Es por ello que la opción preferida es sustituir los valores mayores o iguales a la unidad por 0,999, y los valores menores o iguales a cero por 0,001.
- Se pondera el modelo (4) dividiendo ambos miembros de la ecuación por la desviación típica estimada $\sqrt{\mathbf{s}_i^2} = \sqrt{\hat{Y}_i(1 - \hat{Y}_i)}$, con el fin de transformar el modelo en homocedástico.

$$\frac{Y_i}{\sqrt{\mathbf{s}_i^2}} = \mathbf{b}_1 \frac{1}{\sqrt{\mathbf{s}_i^2}} + \mathbf{b}_2 \frac{X_{li}}{\sqrt{\mathbf{s}_i^2}} + \dots + \mathbf{b}_k \frac{X_{ki}}{\sqrt{\mathbf{s}_i^2}} + \mathbf{e}_i \frac{1}{\sqrt{\mathbf{s}_i^2}} \quad (7)$$

La estimación por MCO del modelo transformado es equivalente a aplicar MCG en el modelo (4) y en ambos casos se obtienen estimaciones eficientes de los coeficientes de regresión.

Sin embargo, uno de los problemas que presenta la estimación por MCG es la pérdida del término independiente en el modelo. La omisión del término independiente puede provocar que la suma de los residuos sea distinta de cero lo que puede tener consecuencias sobre el coeficiente de determinación (puede ser negativo), la función de verosimilitud estimada a partir de los residuos y los estadísticos que se obtienen a partir de ella.

4. El coeficiente de determinación R^2 está subestimado. La suma de los cuadrados de los residuos ($\sum e_i^2$) es más grande de lo habitual debido a la forma específica en que se distribuye la nube de puntos de una variable dicotómica. Dado que el cálculo del coeficiente de determinación² se ve afectado por $\sum e_i^2$, el R^2 calculado en la estimación por MCO es más pequeño de lo que realmente debería ser.

² El coeficiente de determinación se obtiene a través de la fórmula $R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \hat{Y}_i)^2}$

4. MODELOS DE PROBABILIDAD NO LINEAL

La estimación e interpretación de los modelos probabilísticos lineales plantea una serie de problemas que han llevado a la búsqueda de otros modelos alternativos que permitan estimaciones más fiables de las variables dicotómicas. Para evitar que la variable endógena estimada pueda encontrarse fuera del rango (0, 1), las alternativas disponibles son utilizar modelos de probabilidad no lineales, donde la función de especificación utilizada garantice un resultado en la estimación comprendido en el rango 0-1. Las funciones de distribución cumplen este requisito, ya que son funciones continuas que toman valores comprendidos entre 0 y 1.

Especificación de los modelos de elección discreta (Logit y Probit)

Dado que el uso de una función de distribución garantiza que el resultado de la estimación esté acotado entre 0 y 1, en principio las posibles alternativas son varias, siendo las más habituales la función de distribución logística, que ha dado lugar al modelo Logit, y la función de distribución de la normal tipificada, que ha dado lugar al modelo Probit. Tanto los modelos Logit como los Probit relacionan, por tanto, la variable endógena Y_i con las variables explicativas X_{ki} a través de una función de distribución.

En el caso del modelo Logit, la función utilizada es la logística, por lo que la especificación de este tipo de modelos queda como sigue

$$Y_i = \frac{1}{1 + e^{-a - b_k X_{ki}}} + e_i = \frac{e^{a + b_k X_{ki}}}{1 + e^{a + b_k X_{ki}}} + e_i \quad (8)$$

En el caso del modelo Probit la función de distribución utilizada es la de la normal tipificada, con lo que el modelo queda especificado a través de la siguiente expresión,

$$Y_i = \int_{-\infty}^{a + b_k X_{ki}} \frac{1}{(2\pi)^{1/2}} e^{-\frac{s^2}{2}} ds + e_i \quad (9)$$

donde la variable s es una variable “muda” de integración con media cero y varianza uno.

Dada la similitud existente entre las curvas de la normal tipificada y de la logística, los resultados estimados por ambos modelos no difieren mucho entre sí³, siendo las

³ Discrepan, únicamente, en la rapidez con que las curvas se aproximan a los valores extremos, y así la función logística es más achatada que la normal al alcanzar, esta última, más rápidamente los valores extremos (0 y 1).

diferencias operativas, debidas a la complejidad que presenta el cálculo de la función de distribución normal frente a la logística, ya que la primera solo puede calcularse en forma de integral. La menor complejidad de manejo que caracteriza al modelo Logit es lo que ha potenciado su aplicación en la mayoría de los estudios empíricos.

Al igual que en el Modelo Lineal de Probabilidad, el Modelo Logit (8) se puede interpretar en términos probabilísticos, es decir, sirve para medir la probabilidad de que ocurra el acontecimiento objeto de estudio ($Y_i=1$). En cuanto a la interpretación de los parámetros estimados en un modelo Logit, el signo de los mismos indica la dirección en que se mueve la probabilidad cuando aumenta la variable explicativa correspondiente, sin embargo, la cuantía del parámetro no coincide con la magnitud de la variación en la probabilidad (como si ocurría en el MLP). En el caso de los modelos Logit, al suponer una relación no lineal entre las variables explicativas y la probabilidad de ocurrencia del acontecimiento, cuando aumenta en una unidad la variable explicativa los incrementos en la probabilidad no son siempre iguales ya que dependen del nivel original de la misma.

Una interpretación más sencilla del parámetro estimado es la que se obtiene a través de la linealización del modelo. Para ello, partiendo de la ecuación general del Modelo Logit (8) y definido M_i como la probabilidad del estado o la alternativa 1, se tiene

$$E(Y_i) = \text{Prob}(Y_i = 1) = M_i = \frac{e^{a+b_k X_{ki}}}{1 + e^{a+b_k X_{ki}}} \quad (10)$$

de donde:

$$\begin{aligned} M_i + M_i e^{a+b_k X_{ki}} &= e^{a+b_k X_{ki}} \\ M_i &= (1 - M_i) e^{a+b_k X_{ki}} \\ \frac{M_i}{(1 - M_i)} &= e^{a+b_k X_{ki}} \end{aligned} \quad (11)$$

Al cociente entre la probabilidad de que ocurra un hecho, o de que se elija la opción 1, frente a la probabilidad de que no suceda el fenómeno, o de que se elija la opción 0, se la denomina como la ratio odds. Su interpretación es la “ventaja” o preferencia de la opción 1 frente a la 0, es decir, el número de veces que es más probable que ocurra el fenómeno frente a que no ocurra.

$$\text{Ratio odds} = \frac{M_i}{(1 - M_i)}$$

El ratio odds⁴, tal y como está construido (cociente entre probabilidades), siempre será mayor o igual que 0. El campo de variación del ratio va desde 0 hasta $+\infty$, y su interpretación se realiza en función de que el valor sea igual, menor o superior a la unidad: si toma el valor 1 significa que la probabilidad de que ocurra la alternativa 1 es la misma que la de que no ocurra; si el ratio es menor que 1 indica que la ocurrencia de la alternativa 1 tiene menor probabilidad que la ocurrencia de la alternativa 0; mientras que si es mayor que la unidad la opción 1 es más probable que la 0.

El interés de esta medida adquiere sentido cuando se comparan las ventajas para distintos valores de la variable explicativa, calculándose el cociente entre odds. Así, si se compara la situación de la observación “i” con la de la observación “j” (que suele ser la de referencia), el cociente entre odds mide cuanto es más probable que se de la alternativa 1 en “i” que en “j”.

$$\text{Cociente entre odds} = \frac{\frac{M_i}{(1-M_i)}}{\frac{M_j}{(1-M_j)}} = \frac{e^{a+b_k X_{ki}}}{e^{a+b_k X_{kj}}} = e^{b_k (X_{ki} - X_{kj})}$$

Si el valor obtenido es mayor a la unidad, la probabilidad de ocurra la alternativa 1 en la observación “i” es mayor que en la observación “j”, mientras que si el valor obtenido es inferior a uno, la probabilidad de ocurrencia de la alternativa 1 es superior en la observación “j” que en la “i”. Si el valor obtenido es igual a la unidad significa que las probabilidades en ambas observaciones son iguales.

El cálculo del cociente entre odds facilita la interpretación de los parámetros estimados cuando se aplica al caso concreto de calcular la variación en la preferencia o ventaja de un individuo “i” cuando incrementa en una unidad una de las variables explicativas, frente a la ventaja o preferencia del mismo individuo “i” cuando se encuentra en la situación de referencia, obteniéndose para este caso concreto

⁴ Tomando logaritmos neperianos del ratio odds se linealiza la ecuación del modelo Logit, respetando el objetivo de que los valores estimados caigan dentro del rango (0-1), obteniéndose la expresión:

$$\text{Ln}\left(\frac{M_i}{1-M_i}\right) = \text{Ln}(e^{a+b_k X_{ki}}) = a + b_k X_{ki}$$

La nueva variable $\text{Ln}\left(\frac{M_i}{1-M_i}\right)$ generada representa en una escala logarítmica la diferencia entre las probabilidades de que ocurra la alternativa 1 y su contraria.

$$\text{Cociente entre Odds} = \frac{\frac{M_{i+1}}{(1-M_{i+1})}}{\frac{M_i}{(1-M_i)}} = \frac{e^{a+b_k(X_{ki}+1)}}{e^{a+b_k X_{ki}}} = e^{b_k(X_{ki}+1-X_{ki})} = e^{b_k}$$

De donde el parámetro e^{b_k} es un factor de cambio en el cociente entre odds cuando el valor de la variable X_k aumenta en una unidad y el resto de variables explicativas se mantienen constantes. Es decir, el parámetro b_k se interpreta como el número de veces que incrementa el logaritmo de la ventaja o preferencia de la opción 1 frente a la 0 cuando incrementa en una unidad X_k . En muchas ocasiones lo que se analiza es el valor del antilogaritmo del parámetro de tal manera que se evalúe de una forma más directa su efecto sobre la probabilidad.

Estimación de los parámetros en los modelos Logit

Antes de abordar el método de estimación en los modelos Logit, es preciso distinguir la existencia de dos casos diferenciados que implican la utilización de métodos de estimación distintos: los modelos Logit con observaciones repetidas y con observaciones no repetidas.

Para el caso sencillo de una única variable explicativa, nos encontramos en una situación con observaciones repetidas cuando la variable X es discreta y presenta un número reducido de alternativas o intervalos (F), de manera que para cada alternativa de la variable X tendremos n_i observaciones de Y , pudiéndose calcular las proporciones o probabilidades muestrales. En este caso la matriz de n datos muestrales quedará reducida a F observaciones siendo los valores que tome la variable endógena (P_i) las proporciones muestrales calculadas a través de la expresión

$$P_i = \sum_{i=1}^F \frac{Y_i}{n_i} \quad (12)$$

La generalización del modelo a k variables explicativas implica la existencia de observaciones repetidas de Y para cada combinación de las k variables explicativas, pudiéndose calcular las proporciones o probabilidades muestrales de la misma forma que en el caso anterior. En este caso, si bien los valores de la variable endógena están acotados en el rango 0-1, son valores continuos, por lo que el método utilizado para la estimación de los parámetros del modelo es el que habitualmente se utiliza en la econometría tradicional que trabaja con variables continuas.

Por lo tanto, ante la presencia de observaciones repetidas, se podría aplicar el método de Mínimos Cuadrados Ordinarios. Sin embargo, la existencia de heterocedasticidad en el modelo obliga a estimar por Mínimos Cuadrados Generalizados, para garantizar el cumplimiento de las propiedades de los parámetros estimados, utilizándose la inversa de la varianza de los errores como ponderación del modelo.

Sin embargo, lo más habitual es no poder calcular las probabilidades muestrales, bien porque las variables explicativas incluidas en el modelo son continuas, o bien porque aún siendo éstas discretas, la combinación de las mismas impide la obtención de observaciones repetidas de la variable endógena para cada uno de los intervalos F . En esta situación, la matriz de datos muestrales estará formada por n observaciones pudiendo ser el valor de la variable endógena para cada una de ellas 1 ó 0. La naturaleza dicotómica de la variable dependiente en este tipo de modelos impide la utilización de los métodos tradicionales en la estimación de los parámetros, al no poderse calcular la inversa de la varianza utilizada como ponderación del modelo. Para la estimación de los parámetros se utiliza el método de Máxima Verosimilitud.

A continuación se describen ambos métodos de estimación (máxima verosimilitud y mínimos cuadrados generalizados) comenzando por el caso más habitual de ausencia de observaciones repetidas.

A. Estimación con observaciones no repetidas: Método de Máxima-Verosimilitud

Dada una variable aleatoria, caracterizada por unos parámetros, y dada una muestra poblacional, se consideran estimadores Máximo-Verosímiles de los parámetros de una población determinada, aquellos valores de los parámetros que generarían con mayor probabilidad la muestra observada. Es decir, los estimadores Máximo-Verosímiles son aquellos valores para los cuales la función de densidad conjunta (o función de verosimilitud) alcanza un máximo.

Suponiendo que las observaciones son independientes, la función de densidad conjunta de la variable dicotómica Y_i queda como:

$$Prob(Y_1 Y_2 \dots Y_i \dots Y_n) = \prod_{i=1}^n M_i^{Y_i} (1 - M_i)^{1-Y_i} \quad (13)$$

donde M_i recoge la probabilidad de que $Y_i=1$. Por simplicidad se trabaja con la función de densidad conjunta en logaritmos, cuya expresión es:

$$\mathfrak{L} = \ln L = \sum_{i=1}^i Y_i \ln M_i + \sum_{i=1+i}^{n-i} (1 - Y_i) \ln(1 - M_i) = \sum Y_i \ln M_i + \sum (1 - Y_i) \ln(1 - M_i) \quad (14)$$

El método de estimación de máxima verosimilitud elige el estimador del parámetro que maximiza la función de verosimilitud ($\ell = \ln L$), por lo que el procedimiento a seguir será calcular las derivadas de primer orden de esta función con respecto a los parámetros que queremos estimar, igualarlas a 0 y resolver el sistema de ecuaciones resultante. Las derivadas de primer orden de la función de verosimilitud respecto a los parámetros \mathbf{a} y \mathbf{b} , tras pequeñas manipulaciones, quedan como siguen:

$$\frac{\partial \ell}{\partial \mathbf{a}} = \sum_{i=1}^n (Y_i - M_i) = \sum \left(Y_i - \frac{e^{\hat{\mathbf{a}} + \hat{\mathbf{b}}X_i}}{1 + e^{\hat{\mathbf{a}} + \hat{\mathbf{b}}X_i}} \right) = 0 \quad (15)$$

$$\frac{\partial \ell}{\partial \mathbf{b}} = \sum_{i=1}^n (Y_i - M_i)X_i = \sum \left(Y_i - \frac{e^{\hat{\mathbf{a}} + \hat{\mathbf{b}}X_i}}{1 + e^{\hat{\mathbf{a}} + \hat{\mathbf{b}}X_i}} \right) X_i = 0 \quad (16)$$

y sustituyendo M_i por su valor queda:

$$\frac{\partial \ell}{\partial \mathbf{a}} = \sum_{i=1}^n e_i = \sum \left(Y_i - \frac{e^{\hat{\mathbf{a}} + \hat{\mathbf{b}}X_i}}{1 + e^{\hat{\mathbf{a}} + \hat{\mathbf{b}}X_i}} \right) = 0$$

$$\frac{\partial \ell}{\partial \mathbf{b}} = \sum_{i=1}^n X_i e_i = \sum \left(Y_i - \frac{e^{\hat{\mathbf{a}} + \hat{\mathbf{b}}X_i}}{1 + e^{\hat{\mathbf{a}} + \hat{\mathbf{b}}X_i}} \right) X_i = 0$$

Se trata de un sistema de ecuaciones no lineales por lo que es necesario aplicar un método iterativo o algoritmo de optimización que permita la convergencia en los estimadores.

B. Estimación con observaciones repetidas: Método Mínimos Cuadrados Generalizados

La estimación del modelo con datos agrupados podría realizarse mediante el procedimiento habitual utilizado para estimar regresiones lineales, ya que la variable a modelizar ya no es dicotómica (es continua aunque acotada en el rango 0-1). Para ello es necesario linealizar el modelo, lo cual es fácil de realizar a través de la transformación ya comentada anteriormente, y por la cual:

$$\text{Ln} \left(\frac{M_i}{1 - M_i} \right) = \mathbf{a} + \mathbf{b}_k X_{k_i} + \mathbf{e}_i$$

donde \mathbf{e}_i es el valor de la perturbación aleatoria incluida en la especificación de todo modelo de regresión lineal y que cumple las hipótesis de perturbación esférica y ausencia de autocorrelación. El modelo así transformado puede estimarse por el

procedimiento habitual de Mínimos Cuadrados Ordinarios (MCO). Sin embargo, y dado que el valor de M_i es desconocido y debe sustituirse por su estimación muestral P_i , el modelo a estimar quedaría como:

$$\text{Ln}\left(\frac{P_i}{1-P_i}\right) = \mathbf{a} + \mathbf{b}_k X_{ki} + \mathbf{e}_i + \mathbf{e}_i'$$

donde \mathbf{e}_i' recoge el error cometido al utilizar la estimación muestral de la probabilidad P_i , en vez de su valor desconocido M_i . Al sustituir M_i por su estimación muestral P_i , los errores, supuestos independientes, cumplen la condición asintótica de normalidad exigida para realizar contrastaciones y construcción de intervalos de confianza, pero, dejan de cumplir la condición de homocedasticidad ya que su varianza no es constante⁵.

La presencia de heterocedasticidad impide la estimación a través de Mínimos Cuadrados Ordinarios, siendo necesario aplicar el método de Mínimos Cuadrados Generalizados, que sin exigir la condición de homocedasticidad de los errores, permite estimar estimadores ELIO. Este procedimiento transforma el modelo a estimar en otro, donde todas las variables quedan ponderadas por los inversos de las varianzas de los errores, y dado que se desconocen dichos valores verdaderos, éstos se sustituyen por su estimación muestral P_i , de donde:

$$s_i = \frac{1}{\widehat{\text{Var}}(\mathbf{e}_i')} = n_i P_i (1 - P_i) \quad (17)$$

quedando el modelo a estimar como:

$$s_i \text{Ln}\left(\frac{P_i}{1-P_i}\right)_i = \mathbf{a} s_i + \mathbf{b}_k X_{ki} s_i + \mathbf{e}_i \quad (18)$$

Contraste y validación de hipótesis

En el caso de trabajar con observaciones repetidas la contrastación y validación del modelo estimado sigue la misma metodología que la empleada en el análisis de regresión tradicional, por lo que remitimos a éste para profundizar en este tema. Mientras que si nos encontramos en el caso de no disponer de observaciones repetidas,

⁵ La varianza de la perturbación aleatoria no es homocedástica ya que depende del nivel en que se encuentre la variable explicativa X, al definirse

$$\mathbf{e}_i' = \text{Ln}\left(\frac{P_i}{1-P_i}\right) - \text{Ln}\left(\frac{M_i}{1-M_i}\right)$$

la etapa de contrastación y validación del modelo estimado por máxima-verosimilitud se lleva a cabo aplicando los estadísticos específicos que se comentan a continuación.

A. Significatividad estadística de los parámetros estimados

La distribución del estimador del parámetro \mathbf{b} es aproximadamente:

$$N\left(\mathbf{b}; \sqrt{\text{Var}(\hat{\mathbf{b}})}\right)$$

En tal situación, se puede construir un intervalo de confianza del parámetro estimado, para testar si dicho valor es significativamente distinto de cero de forma individual. El contraste a realizar quedaría definido como:

$$\begin{aligned} H_0 : \mathbf{b} = 0 & \quad \text{El parámetro es igual a cero} \\ H_1 : \mathbf{b} \neq 0 & \quad \text{El parámetro es distinto de cero} \end{aligned}$$

El intervalo de confianza proporciona un rango de posibles valores para el parámetro, por lo que si el valor estimado no pertenece a dicho intervalo, se deberá rechazar la hipótesis nula. El intervalo quedaría definido como:

$$\hat{\mathbf{b}} - z_{\alpha/2} \sqrt{\text{Var}(\hat{\mathbf{b}})} \leq \mathbf{b} \leq \hat{\mathbf{b}} + z_{\alpha/2} \sqrt{\text{Var}(\hat{\mathbf{b}})}$$

donde α es la probabilidad de que el verdadero valor del parámetro \mathbf{b} se halle fuera del intervalo, y z es el valor tabular de la distribución $N(0;1)$ que deja a su derecha una probabilidad igual a $\alpha/2$.

A partir de la expresión anterior se puede fijar un rechazo de la hipótesis nula cuando:

$$\left| \frac{\hat{\mathbf{b}}}{\sqrt{\text{Var}(\hat{\mathbf{b}})}} \right| \geq z_{\alpha/2}$$

B. Medidas de bondad de ajuste del modelo

El uso de la función de verosimilitud en la estimación, hace que la bondad del ajuste en los modelos de elección discreta sea un tema controvertido, ya que en estos modelos no existe una interpretación tan intuitiva como en el modelo de regresión clásico. A continuación se describen los contrastes más utilizados en la literatura econométrica para medir la bondad de ajuste en un modelo Logit y que concretaremos en: índice de

cociente de verosimilitudes, el estadístico chi-cuadrado de Pearson, el porcentaje de aciertos estimados en el modelo, y la prueba de Hosmer-Lemeshow.

B.1. Índice de cociente de verosimilitudes

La función de verosimilitud puede también utilizarse para obtener un estadístico, que tiene cierta semejanza con el coeficiente de determinación calculado en la estimación lineal, conocido “índice de cociente de verosimilitudes”. Este estadístico compara el valor de la función de verosimilitud de dos modelos: uno corresponde al modelo estimado que incluye todas las variables explicativas (modelo completo) y el otro sería el del modelo cuya única variable explicativa es la constante (modelo restringido). El estadístico, también conocido como R^2 de McFadden ya que fue propuesto por McFadden en 1974, se define como:

$$RV = ICV = 1 - \frac{\log L}{\log L(0)} \quad (19)$$

donde L es el valor de la función de verosimilitud del modelo completo (el estimado con todas las variables explicativas) y $L(0)$ es el valor correspondiente del modelo restringido (el que incluye únicamente en la estimación el término constante).

El ratio calculado tendrá valores comprendidos entre 0 y 1 de forma que:

- Valores próximos a 0 se obtendrán cuando $L(0)$ sea muy parecido a L , situación en la que nos encontraremos cuando las variables incluidas en el modelo sean poco significativas, es decir, la estimación de los parámetros \mathbf{b} no mejora el error que se comete si dichos parámetros se igualaran a 0. Por lo que en este caso la capacidad explicativa del modelo será muy reducida.
- Cuanto mayor sea la capacidad explicativa del modelo, mayor será el valor de L sobre el valor de $L(0)$, y más se aproximará el ratio de verosimilitud calculado al valor 1.

B.2. Una medida del error: el estadístico χ^2 de Pearson

Para medir la bondad del ajuste también se utilizan medidas del error que cuantifican la diferencia entre el valor observado y el estimado. En concreto, para contrastar la hipótesis nula de que

$$H_0 : Y_i = \hat{M}_i; \text{ lo que equivale a } H_0 : Y_i - \hat{M}_i = e_i = 0$$

se construye un estadístico que recoge los residuos estandarizados o de Pearson⁶ del modelo Logit, que se definen como la diferencia entre el valor observado de la variable respuesta y el estimado, dividido por la estimación de la desviación típica, ya que la esperanza es nula. A través del contraste de multiplicadores de Lagrange, se puede calcular el estadístico conocido con el nombre de c^2 de Pearson, que se define como

$$c^2 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n \frac{(Y_i - \hat{M}_i)^2}{\hat{M}_i(1 - \hat{M}_i)} \quad (20)$$

Este estadístico es similar a la suma de cuadrados de los residuos del modelo de regresión convencional. El ajuste del modelo será mejor cuanto más cerca esté el valor del estadístico de cero. Para saber a partir de que valor puede considerarse el ajuste como aceptable es necesario conocer la distribución del estadístico. Éste estadístico, bajo la hipótesis nula, se distribuye como una chi-cuadrado con $(n-k)$ grados de libertad, por lo que su valor se compara con el valor teórico de las tablas de la chi-cuadrado para contrastar la hipótesis nula. Si el valor calculado es superior al valor teórico se rechaza la hipótesis nula lo que equivale a decir que el error cometido es significativamente distinto de cero, es decir, se trataría de un mal ajuste.

B.3 Porcentaje de aciertos estimados en el modelo

Otra de las vías utilizadas para determinar la bondad de un modelo Logit es predecir con el modelo los valores de la variable endógena Y_i de tal manera que $Y_i = 1$ si $\hat{M}_i > c$ ó $Y_i = 0$ si $\hat{M}_i < c$. Generalmente, el valor que se asigna a c para determinar si el valor de la predicción es igual a 1 o a 0 es de 0,5, puesto que parece lógico que la predicción sea 1 cuando el modelo dice que es más probable obtener un 1 que un 0.

Sin embargo, la elección de un umbral igual a 0,5 no siempre es la mejor alternativa. En el caso en que la muestra presente desequilibrios entre el número de unos y el de ceros la elección de un umbral igual a 0,5 podría conducir a no predecir ningún uno o ningún cero. Así, supuesta una muestra de 1.000 observaciones donde 100 son 1 y el resto 0, si el modelo incluye término constante, la media de las probabilidades estimadas en la muestra será 0,1⁷, por lo que será casi imposible que se obtenga un valor estimado superior a 0,5. Si el umbral seleccionado es de 0,5, con esta regla nunca se llegarían a

⁶ Los residuos estandarizados o de Pearson se definen como:

$$= \frac{Y_i - \hat{M}_i}{\sqrt{\hat{M}_i(1 - \hat{M}_i)}}$$

⁷ Como se ha comentado anteriormente, de la condición de primer orden que queda recogida en la ecuación (III.44) se deduce que la media de las probabilidades estimadas por el modelo, ha de coincidir con la proporción de unos que haya en la muestra.

estimar valores iguales a 1. El modo de resolver este problema es tomar un umbral más pequeño.

Con cualquier tipo de regla predictiva similar se cometerán dos errores: habrá ceros que se clasifiquen incorrectamente como unos y unos que se clasifiquen incorrectamente como ceros. Si se reduce el umbral por debajo de 0,5 aumentará el número de veces que se clasifican correctamente observaciones para las que $Y_i=1$, pero también aumentará el número de veces en que se clasifiquen observaciones como unos para las que $Y_i=0$. Cambiando el valor del umbral se reducirá siempre la probabilidad de un error de un tipo y se aumentará la probabilidad del otro tipo de error. Por lo que el valor que debe tomar el umbral depende de la distribución de datos en la muestra y de la importancia relativa de cada tipo de error.

Una vez seleccionado el nivel del umbral, y dado que los valores reales de Y_i son conocidos, basta con contabilizar el porcentaje de aciertos para decir si la bondad del ajuste es elevada o no. A partir de este recuento se puede construir el siguiente cuadro de clasificación:

Cuadro de clasificación de aciertos

		Valor real de Y_i	
		$Y_i = 0$	$Y_i = 1$
Predicción de \hat{M}_i	$\hat{M}_i < c$	P_{11}	P_{12}
	$\hat{M}_i > c$	P_{21}	P_{22}

Donde P_{11} y P_{22} corresponderán a predicciones correctas (valores 0 bien predichos en el primer caso y valores 1 bien predichos en el segundo caso), mientras que P_{12} y P_{21} corresponderán a predicciones erróneas (valores 1 mal predichos en el primer caso y valores 0 mal predichos en el segundo caso). A partir de estos valores se pueden definir los índices que aparecen en el siguiente cuadro.

Índices para medir la bondad del ajuste

Índice	Definición	Expresión
Tasa de aciertos	Cociente entre las predicciones correctas y el total de predicciones	$\frac{P_{11} + P_{22}}{P_{11} + P_{12} + P_{21} + P_{22}}$
Tasa de errores	Cociente entre las predicciones incorrectas y el total de predicciones	$\frac{P_{12} + P_{21}}{P_{11} + P_{12} + P_{21} + P_{22}}$
Especificidad	Proporción entre la frecuencia de valores 0 correctos y el total de valores 0 observados	$\frac{P_{11}}{P_{11} + P_{21}}$
Sensibilidad	Razón entre los valores 1 correctos y el total de valores 1 observados	$\frac{P_{22}}{P_{12} + P_{22}}$
Tasa de falsos ceros	Proporción entre la frecuencia de valores 0 incorrectos y el total de valores 0 observados	$\frac{P_{21}}{P_{11} + P_{21}}$
Tasa de falsos unos	Razón entre los valores 1 incorrectos y el total de valores 1 observados	$\frac{P_{12}}{P_{12} + P_{22}}$

B.4. Prueba de Hosmer-Lemeshow

Otra medida global de la exactitud predictiva, no basada en el valor de la función de verosimilitud sino en la predicción real de la variable dependiente, es el contraste de clasificación diseñado por David W. Jr. Hosmer y Stanley Lemeshow en 1989. Dicho contraste consiste en realizar comparaciones entre el valor estimado y el observado por grupos. Para ello las observaciones se dividen en J grupos (generalmente 10) aproximadamente iguales, dividiendo el recorrido de la probabilidad en deciles de riesgo (esto es probabilidad de ocurrencia del fenómeno < 0.1 , < 0.2 , y así hasta < 1). Cada uno de los grupos contiene n_j observaciones, y en cada uno de los J grupos se define:

- Y_j como la suma de los valores 1 en cada uno de los grupos ($Y_j = \sum Y_i$)
- \bar{P}_j como la media de los valores predichos en cada grupo ($\bar{P}_j = \sum \frac{\hat{P}_i}{n_j}$).

A partir de esta información se puede construir una tabla de contingencia a través de la que se compara tanto la distribución de ocurrencia, como la de no ocurrencia prevista por la ecuación y los valores realmente observados. El contraste se realiza comparando las frecuencias observadas y esperadas a través del cálculo del estadístico

$$HL = \sum_{j=1}^J \frac{(Y_j - n_j \bar{P}_j)^2}{n_j \bar{P}_j (1 - \bar{P}_j)} \quad (21)$$

Hosmer y Lemeshow demuestran que cuando el modelo es correcto el estadístico HL sigue una distribución chi-cuadrado con $J-2$ grados de libertad, por lo que valores inferiores del estadístico calculado respecto al teórico indicarán un buen ajuste del modelo.

El uso correcto de este contraste requiere un tamaño de muestra adecuado para asegurar que cada grupo cuenta al menos con cinco observaciones. Además el estadístico chi-cuadrado es sensible al tamaño muestral, permitiendo que esta medida encuentre diferencias estadísticamente muy pequeñas cuando el tamaño muestral crece.

5. MODELOS DE RESPUESTA MÚLTIPLE

Cuando la variable endógena a modelizar es una variable discreta con varias alternativas posibles de respuesta (J) nos encontramos ante los modelos de respuesta múltiple. Estos modelos se clasifican en dos grandes grupos según que las alternativas que presenta la variable endógena se puedan ordenar (modelos con datos ordenados) o no se puedan ordenar (modelos con datos no ordenados).

Modelos de respuesta múltiple con datos no ordenados

La especificación general de los modelos de respuesta múltiple con datos no ordenados queda recogida a través de la siguiente expresión:

$$\text{Pr } ob(Y_i = j) = \frac{e^{b'Z_{ij}}}{\sum_{j=0}^J e^{b'Z_{ij}}} \quad (22)$$

donde Z_{ij} representa la matriz de los regresores del modelo. Dichas variables explicativas pueden ser de dos tipos:

- Variables que contienen aspectos específicos del individuo y por tanto, su valor será el mismo en todas las alternativas. Este tipo de variables reciben el nombre de características, y se las denota por W_i .
- Variables que contienen aspectos específicos de las alternativas entre las que se ha de elegir, y varían tanto entre individuos como entre alternativas. Este tipo de variables reciben el nombre de atributos de las alternativas y se las denota por X_{ij} .

A partir de esta especificación general, y teniendo en cuenta que la inclusión en el modelo de variables explicativas que hagan referencia a características o atributos

permite la especificación de modelos diferentes denominados, modelo logit multinomial en el primer caso y modelo logit condicional en el segundo.

A. El modelo Logit Multinomial

Este tipo de modelos es el que se utiliza con más frecuencia en los trabajos aplicados. En este modelo los valores de las variables explicativas varían para cada individuo pero son constantes para cualquier alternativa, por lo que no se puede apreciar la influencia de la variable en cada alternativa a no ser que se introduzca una variable ficticia, multiplicada por los valores de W_i , que represente a cada alternativa. Para evitar problemas de singularidad, el número de variables ficticias a introducir en el modelo será igual al número de alternativas menos uno ($J-1$).

La formulación de un Logit Multinomial queda recogida a través de la siguiente ecuación⁸:

$$\Pr ob(Y_i = j) = P_{ij} = \frac{e^{b_j X_i}}{\sum_{j=0}^{J-1} e^{b_j X_i}} \quad (23)$$

donde j representa el índice asociado a cada alternativa y va desde 0 hasta ($J-1$). El vector de parámetros lleva asociado el subíndice correspondiente a la alternativa concreta analizada. Las ecuaciones estimadas proporcionan un conjunto de probabilidades para cada una de las alternativas que puede tomar un individuo i y tenga X_i como características individuales.

En el modelo Logit Multinomial existe una indeterminación cuando se trata de estimar el valor de los parámetros. Para solucionar este problema se normaliza el modelo tomando para los parámetros que acompañan a la alternativa cero el valor cero, $b_0 = 0$.

Las probabilidades resultantes son

$$\Pr ob(Y_i = j) = \frac{e^{b_j X_i}}{1 + \sum_{j=1}^{J-1} e^{b_j X_i}} \quad \text{para } j=1, 2, \dots, (J-1) \quad (24)$$

⁸ A pesar de que las características específicas de cada individuo se han denotado con W_i , en lo que sigue se denotarán con X_i al hacer referencia a las variables explicativas de un modelo econométrico en el que tradicionalmente se utiliza esa denominación.

$$\text{Pr } ob(Y_i = 0) = \frac{1}{1 + \sum_{j=1}^{J-1} e^{b_j X_i}} \quad \text{para } j=0$$

Donde se tiene que cumplir que

$$\sum_{j=0}^{J-1} P_j = 1$$

Para el caso sencillo de un modelo en el que la variable endógena presenta tres posibles alternativas de elección y sólo existe una variable explicativa en la modelización, la probabilidad asociada a cada una de las alternativas posibles de elección tomarían las siguientes expresiones⁹

$$\begin{aligned} P_0 &= \frac{1}{1 + e^{a_1 + b_1 X_i} + e^{a_2 + b_2 X_i}} \\ P_1 &= \frac{e^{a_1 + b_1 X_i}}{1 + e^{a_1 + b_1 X_i} + e^{a_2 + b_2 X_i}} \\ P_2 &= \frac{e^{a_2 + b_2 X_i}}{1 + e^{a_1 + b_1 X_i} + e^{a_2 + b_2 X_i}} \end{aligned} \quad (25)$$

con $P_0 + P_1 + P_2 = 1$

Y la matriz de diseño X vendrá expresada como

$$X = \begin{bmatrix} 1 & 0 & X_1 & 0 \\ 0 & 1 & 0 & X_1 \\ 1 & 0 & X_2 & 0 \\ 0 & 1 & 0 & X_2 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & X_n & 0 \\ 0 & 1 & 0 & X_n \end{bmatrix}$$

B. El modelo Logit Condicional

Cuando las variables explicativas que se utilizan para estimar las probabilidades asociadas a cada una de las posibles alternativas que presenta la variable endógena se

⁹ Se identifican con la letra **a** los parámetros que acompañan al término independiente y con la letra **b** los que acompañan a las variables explicativas.

refieren a atributos de las distintas alternativas, y no a características específicas de los individuos, el modelo que se utiliza en la estimación es el llamado Logit Condicional. En este caso, el valor de cada variable variará para cada alternativa y puede hacerlo o no para cada individuo.

La diferencia de este modelo con el Logit Multinomial es que en este caso solo existe un vector de parámetros a estimar, mientras que en el caso anterior existían tantos vectores como alternativas menos una. Es por ello, que en la formulación del modelo el vector de parámetros, al ser único, no lleva asociado ningún subíndice relacionado con la alternativa a la que acompaña, como ocurría en el caso anterior.

La otra diferencia hace referencia a que en este caso no existe ninguna indeterminación a la hora de estimar los parámetros, por lo que no es necesario igualar ningún vector \mathbf{b} a cero.

La expresión formal del modelo queda definida como

$$\Pr ob(Y_i = j) = \frac{e^{\mathbf{b}'X_{ij}}}{\sum_{j=1}^J e^{\mathbf{b}'X_{ij}}} \quad \text{para } j = 1, 2, \dots, J \quad (26)$$

C. El modelo Logit Anidado

Uno de los problemas que se plantean en los modelos expuestos de respuesta múltiple es el de que se construyen bajo la hipótesis de presencia de alternativas irrelevantes o superfluas, según la cual la relación entre las probabilidades de decidir entre dos alternativas no depende del resto de las alternativas. Esta propiedad se debe al supuesto inicial de que las perturbaciones aleatorias del modelo son independientes, es decir, las perturbaciones afectan de la misma forma a la diferencia de utilidad entre cualquier par de alternativas. El caso contrario sería la presencia de autocorrelación en el modelo, lo cual se daría, por ejemplo, cuando un individuo percibe unas alternativas más similares entre sí que otras.

Si bien asumir la hipótesis de independencia de las alternativas irrelevantes simplifica el proceso de estimación, supone una restricción en la modelización del comportamiento de los individuos que no parece razonable en determinadas circunstancias. Así, esta propiedad carece de validez cuando algunas de las alternativas son sustitutivos cercanos, ya que en este caso existirían alternativas correlacionadas. Como alternativa para relajar la hipótesis de independencia de alternativas irrelevantes, se ha desarrollado el modelo Logit anidado o Logit jerárquico (que en terminología anglosajona es conocido como Nested Logit).

La construcción del modelo se realiza agrupando el conjunto de alternativas posibles en subgrupos y manteniendo la hipótesis de independencia de alternativas irrelevantes dentro de cada grupo y en la elección entre grupos. En este modelo, la elección de una de las alternativas posibles se realiza en dos o más etapas, definiéndose una estructura arbórea: primero se escoge entre los conjuntos de alternativas y después se elige una alternativa específica perteneciente al conjunto seleccionado en principio.

Suponiendo que las J alternativas posibles pueden dividirse en L conjuntos de alternativas, y que las variables explicativas del modelo son $X_{j/l}$, las que se relacionan con las alternativas dentro de un grupo, y Z_l , las que se relacionan con los conjuntos de alternativas, la forma matemática del modelo queda expresada como:

$$P_{j/l} = \frac{e^{b'X_{j/l}}}{\sum_{j=1}^{J_l} e^{b'X_{j/l}}} \quad (27)$$

$$P_l = \frac{e^{g'Z_l + t_l I_l}}{\sum_{j=1}^{J_l} e^{g'Z_l + t_l I_l}}$$

donde $I_l = \ln \sum_{j=1}^{J_l} e^{b'X_{j/l}}$.

Uno de los aspectos problemáticos de este modelo radica en la especificación de la estructura arbórea. En algunos casos, la partición en subgrupos del conjunto de alternativas posibles se hace de modo natural. Sin embargo, en otros casos, dicha partición del conjunto de posibles alternativas se hace sin ningún criterio lógico, por lo que resulta preocupante que los resultados obtenidos dependan de cómo se han definido las ramas. De momento, no existe ningún contraste que permita seleccionar la mejor estructura arbórea de entre varias, por lo que muchos de los trabajos empíricos que estiman este modelo presentan los resultados supuestas distintas especificaciones de la estructura arbórea.

Modelos de respuesta múltiple con datos ordenados

Cuando la variable dependiente es discreta, pero sus valores indican un orden, no es correcto realizar la estimación de la misma a través de los modelos presentados en el apartado anterior, ya que la inclusión de la información que aporta el orden de las alternativas en la especificación del modelo permite obtener unos mejores resultados. Tampoco sería correcto el uso de un modelo de regresión clásico, ya que codificadas las posibles alternativas como 0, 1, 2, ..., $(j+1)$, ..., J , se estaría considerando la diferencia entre $(j+1)$ y $(j+2)$ como la existente entre 1 y 2, lo cual no tiene porque ser así ya que

los números utilizados en la codificación solo representan un orden dentro de una clasificación.

La formulación del modelo Logit ordenado queda como sigue:

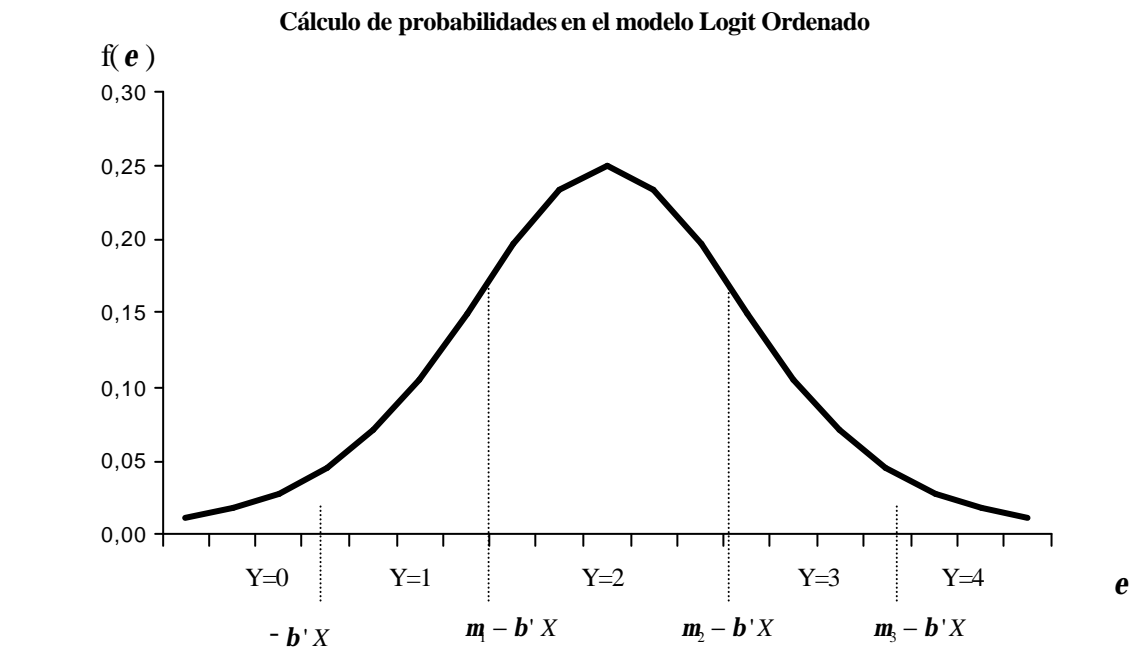
$$\begin{aligned}
 Prob(Y_i = 0) &= \Lambda(-\mathbf{b}' X_i) \\
 Prob(Y_i = 1) &= \Lambda(\mathbf{m}_1 - \mathbf{b}' X_i) - \Lambda(-\mathbf{b}' X_i) \\
 Prob(Y_i = 2) &= \Lambda(\mathbf{m}_2 - \mathbf{b}' X_i) - \Lambda(\mathbf{m}_1 - \mathbf{b}' X_i) \\
 &\dots \\
 Prob(Y_i = (J - 1)) &= 1 - \Lambda(\mathbf{m}_{(J-2)} - \mathbf{b}' X_i)
 \end{aligned} \tag{28}$$

donde $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{(J-2)}$ son parámetros que representan los valores de los umbrales o barreras y se estiman a la vez que \mathbf{b} y $\Lambda(\mathbf{b}' X_i)$ representa la función de distribución logística¹⁰.

Para que todas las probabilidades sean positivas se debe cumplir

$$0 < \mathbf{m}_1 < \mathbf{m}_2 < \dots < \mathbf{m}_{(J-2)}$$

El siguiente gráfico, para el que la variable observada presenta cinco posibles alternativas, sirve para ilustrar la estructura que subyace en la construcción del modelo Logit ordenado.



¹⁰ $\Lambda(\mathbf{b}' X_i) = \frac{e^{\mathbf{b}' X_i}}{1 + e^{\mathbf{b}' X_i}}$